Michael Stengel
stengel@cg.tu-bs.de
Computer Graphics Lab, TU Braunschweig

Dr. Ing. Martin Eisemann
eisemann@cg.tu-bs.de
Computer Graphics Lab, TU Braunschweig

Stephan Wenger
wenger@cg.tu-bs.de
Computer Graphics Lab, TU Braunschweig

Benjamin Hell
hell@cg.tu-bs.de
Computer Graphics Lab, TU Braunschweig

Prof. Dr. Ing. Marcus Magnor
magnor@cg.tu-bs.de
Computer Graphics Lab, TU Braunschweig

Optimizing Apparent Display Resolution Enhancement for Arbitrary Videos

Technical Report 2013-2-18

February 5, 2013

Computer Graphics Lab, TU Braunschweig

Contents

Abstract

Display resolution is frequently exceeded by available image resolution. Recently, apparent display resolution enhancement techniques (ADRE) have demonstrated how characteristics of the human visual system can be exploited to provide super-resolution on high refresh rate displays. In this paper we address the problem of generalizing the apparent display resolution enhancement technique to conventional videos of arbitrary content. We propose an optimization-based approach to continuously translate the video frames in such a way that the added motion enables apparent resolution enhancement for the salient image region. The optimization takes the optimal velocity, smoothness and similarity into account to compute an appropriate trajectory. Additionally, we provide an intuitive user interface which allows to guide the algorithm interactively and preserve important compositions within the video. We present a user study evaluating apparent rendering quality and demonstrate versatility of our method on a variety of general test scenes.

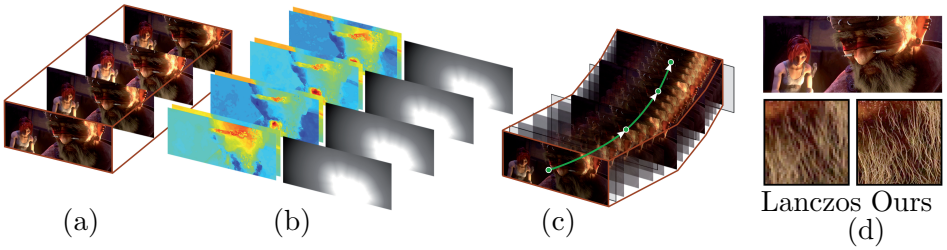

Figure 1: Overview of our proposed technique to improve apparent display resolution enhancement of general video footage. (a) Given a standard 24–30 Hz video we compute (b) optical flow and importance maps to (c) temporally upsample and offset the video along a smooth trajectory. (d) This results in an increased perceptual resolution on the retina that exceeds the physical resolution of the display when downsampled using apparent display resolution enhancement algorithms in comparison to other downsampling schemes.

1 Introduction

Modern cameras and rendering hardware are able to produce highly detailed images. Sophisticated tone and gamut mapping algorithms adapt them to the available display capabilities. Even though hardware constantly evolves, limitations in color, luminance, and spatial resolution constrain the range of reproducible images on various devices. Latest advancements such as apparent image contrast [1] or apparent brightness [2] have shown that it is possible to go beyond the physical limitations of display devices by exploiting characteristics of the human visual system (HVS).

This work addresses the problem of apparent spatial resolution enhancement. High-definition TVs and projectors have already become ubiquitous, but the resolution of current digital cameras and cinema movies is up to one order of magnitude higher than these displays can currently show. The necessary downsampling procedure results in the loss of fine details such as fur, hair or general high-frequency image features. On the other hand, the refresh rate of commodity TVs and projectors increases more and more and 120 Hz TVs are available today. With active-matrix organic light-emitting diode (AMOLED) technology even higher refresh rates ($> 1000\,\mathrm{Hz}$) can be achieved and will be available in the near future. The challenge is how to provide a better viewing experience given the available high-resolution data and the limited hardware resolution. It has been shown by Didyk *et al.* that the integration on the retina of moving high frame rate low-resolution subimages results in an increased perceived resolution, if displayed above the *critical flicker frequency* [3][4]. The necessary smooth pursuit eye movement (SPEM) was induced by artificially moving a static image at constant velocity. Templin *et al.* have shown that a similar effect can be achieved when

exploiting the natural movement in high frame rate videos [5].

The effect of the apparent display resolution enhancement (ADRE) technique is affected by several aspects: first, for a high perceived contrast high refresh rates are necessary [4]. Second, for best perceived spatial resolution the movement in the displayed video needs to be along one of the four diagonals and at a specific velocity [3]. The more the movement differs from these requirements the less noticeable the effect will be.

The approach described in this article (Fig. 1) extends the work of Didyk *et al.* [3] and Templin *et al.* [5] in several important aspects. We show how slight changes to a standard high-resolution, low frame rate video can support the ADRE for a higher perceived resolution. We compute the flow of the most salient regions in the video and shift the video content to enforce a stronger diagonal movement at the required speed. We pay specific attention to subtle and smooth changes which incorporate the natural movements in the video. Our optimization is based on an energy minimization which incorporates saliency (where does a viewer usually look at in the video ?), smoothness (to support SPEM and prevent flickering), similarity to the original footage (to prevent the movement from going astray) and resemblance to the optimal direction and velocity (to provide the best possible input to the ADRE algorithm). In addition, to handle low frame rate videos, a motion path is computed to offset duplicated frames to further support apparent resolution enhancement. A specialized user interface allows to interactively change the optimization parameters within the video for artistic guidance of the optimization. In contrast to [5] and [6], our approach enables apparent resolution enhancement even for scenes that do not contain any movement and for which typical optical flow computations are difficult or impossible. We encourage the reader to check the accompanying video and demo application to get a better impression of the resolution enhancement effect.

One possible application for our approach are common high refresh rate TVs and projectors, for which our approach can be used to display videos containing arbitrary movement perceived at an increased resolution. Furthermore, in mobile phones, cheaper low resolution screens could be used instead of expensive Retina Displays™ to reproduce videos. When watching a video on a display with a different aspect ratio, e.g. 4:3 instead of 16:9, the cropping area can be optimized to support ADRE.

The remainder of this paper is structured as follows: We first present related work in Section 2. In Section 3 we provide background information on the general apparent resolution model and describe its inherent problems for traditional videos in Section 4. We then present our extended model in Section 5 and our two-stage saliency scheme in Section 6. The application of this extended model provides important input to our trajectory optimization in Section 7. In Section 8 we describe a user interface that provides means to preserve certain artistic camera motions or manually restrict the

optimization. Our perceptual experiments and user studies are analyzed in Section 9, before we conclude in Section 10 and 11.

2 Related Work

Displaying high-resolution images on a low-resolution display is a sampling problem. For reconstruction, the high-resolution image is convolved with a reconstruction filter for every output pixel. Well known examples are the cubic splines derived by Mitchell and Netravali [7] and the Lanczos filter. In this work we are interested in approaches that rely on actively adapting the spatial and temporal signal integration to go beyond the physical pixel resolution.

Specialized Hardware Digital light processing video projectors rely on the temporal integration of the human eye by displaying the *RGB* color components sequentially at refresh rates above the *critical flicker frequency*. A disadvantage of these approaches are color falsifications ("rainbow effect") on saccadic eye movements. Berthouzoz and Fattal [8] exploit this integration effect to achieve resolution enhancement by small-amplitude vibrations to a display, synchronized with the screen refresh cycles. Damera-Venkata and Chang [9] combine multiple display samples via the superimposition of image subframes from multiple projectors. This *display supersampling* can be optimized for antialiased and perceived super-resolution images. Wobulated projectors use an opto-mechanical image shifter to slightly shift subimages for this purpose [10]. These approaches require very specialized or calibrated hardware and are not applicable to general devices, wheras we only require a 120 Hz display.

Perceptual resolution enhancement Hara and Shiramatsu [11] inspected the influence of special pixel-color mosaics when moving an image at a specific velocity across the display but could not observe any improvement for the standard *RGB* layout. Similarily, *subpixel rendering* exploits knowledge about the arrangement of *RGB* color filters on a grid of photosensors for optimal filtering [12] or masking defective subpixels [13]. Our work is closely connected to the perceptual approaches by Didyk *et al.* [3] and Templin *et al.* [5]. Both take the HVS's SPEM into account to display subimages at high-refresh rates. The temporal integration in the human eye provides perceived resolution enhancement. A similar approach was taken by Basu and Baudisch [14] who proposed to move the image on a small circular path which has proven non-optimal [5]. While Didyk *et al.* [3] demonstrated the applicability of their approach only for linear motion, Templin *et al.* [5] transformed it into an optimization problem which could be used for arbitrary motion in animations. Fattal *et al.* combine this approach with a super-resolution

method to spatially enhance input videos that are at the display resolution [6] and formulate a perceived video model used in the optimization. Compared to this method in our approach we work with high-resolution videos as input to best possibly keep the originally available details. A necessary requirement for the resolution enhancement is knowledge about the optical flow between the displayed frames. We extend current approaches to handle even difficult scenes containing problematic motion where optical flow might fail, is non-existent or suboptimal for resolution enhancement.

Visual Attention In [5] and [6] optical flow is used for optimizing ADRE. However, movement is not the only attractor for visual attention. Saliency models provide a better approach to predict eye fixations of an observer. Two general approaches exist. Either low-level features like color, intensity, etc. are employed [15] or higher level semantic knowledge including face detection [16] or person detection [17] can be integrated. Cerf *et al.* [18] combine both in a hybrid approach. [19] formulate an focus of attention model for videos. Based on low-level features and a motion activity map the model describes how viewers track a single region of interest in the video frame after quickly scanning the screen.

3 Previous Model

The model of Didyk *et al.* describes the response r of a receptor in the human eye as an integral of the observed intensities I over a time T [3, 4]. When an observer focuses on a detail in a moving image or video, the eye will try to follow its trajectory in a smooth pursuit eye motion. If the receptor moves on a smooth path $p(t)$ the integrated result is:

$$r(I, p(.)) = \int_0^T I(p(t), t)dt \tag{1}$$

Thus, intensities of neigboring pixels are perceptually mixed if the path crosses a boundary. Owing to this movement and the higher density of photoreceptors on the retina in comparison to the screen resolution, neighboring receptors may reveal a different solution to the integral (*hold-type blur*). Equation (1) does not hold in general [20] although it is a valid assumption for signals displayed above the *critical flicker frequency* as subpixel intensities are fused to a steady appearance [4]. Since I is a discrete function in space (pixels) and time (frames), Equation (1) can be reformulated as

$$r(I, p(.)) = \int_0^T I(p(t), t)dt = \sum_{i,j,k} w_{i,j,k} I^k_{i,j} \quad , \tag{2}$$

where

$$w_{i,j,k} = \int \chi_{i,j}(p(t))\chi_k(t)dt \quad . \tag{3}$$

The characteristic function $\chi_{i,j}(p(t))$ equals one if $p(t)$ lies within the pixel (i, j) and zero otherwise; $\chi_k(t)$ is a similar function for time, that means frames. The weight $w_{i,j,k}$ is normalized by the total length of the path $|p|$. Utilizing the hold-type blur induced by the SPEM in combination with high refresh rate screens results in the apparent display resolution enhancement [3]. The intent is to optimize the subimages so that

$$\mathbf{W} \begin{pmatrix} I_L^1 \\ \vdots \\ I_L^k \end{pmatrix} - I_H = 0 \quad , \tag{4}$$

where I_L^i is the i-th low-resolution subimage and I_H is the original high-resolution image. The underlying assumption is that there is a one-to-one mapping between receptors and pixels in the high-resolution image so that $r_{x,y}$ is close to $I_H(x, y)$, i.e. one row in $\mathbf{W}$ describes the path of one receptor along the subimages. If the subframes are displayed at high refresh rates, integration of intensities in the retina can reconstruct the high frequency details because of SPEM.

Templin *et al.* [5] extend this model to videos by approximating the complex motion in an animation with many simple integer motions computed for every possible triplet of frames in an animation, i.e. for subframes {1,2,3}, {2,3,4}, etc.

$$\mathbf{W} \begin{pmatrix} \mathbf{I_L} \end{pmatrix} - \mathbf{I_H} = 0, \tag{5}$$

where $\mathbf{I_L}$ is the vector of all subframes and $\mathbf{I_H}$ the vector of the original high-resolution images. The triplets are encoded in the appropriate weighting matrix $\mathbf{W}$.

4 Problem Statement

Unfortunately, for natural videos no sufficiently accurate solution to Equation (5) may exist. Between eye saccades the foveal area of the eye follows a focused feature f in a video. In that case the path $p(t)$ is dependent on the movement of f. Several cases exist where ADRE fails, Fig. 2. If f does not move at all or is too slow (r_1, orange), the integral of neighboring receptors on the same low-resolution pixel exhibit the same result for Equation (2). Thus, no resolution enhancement is possible. If f moves faster than $2.5\,{}^{\mathrm{deg}}/_{\mathrm{s}}$ (r_2, blue), stabilization of the image on the retina cannot be guaranteed anymore [21]. In the case of horizontal or vertical movement (r_3, red), resolution enhancement is only possible in the same direction. Owing to reaction times of the HVS the eye cannot follow sudden kinks in the movement of f (r_4, green). Optimal apparent display resolution enhancement is achieved only if f moves along the diagonal at the speed of one high-resolution pixel per subframe in x and y direction (r_5, magenta).

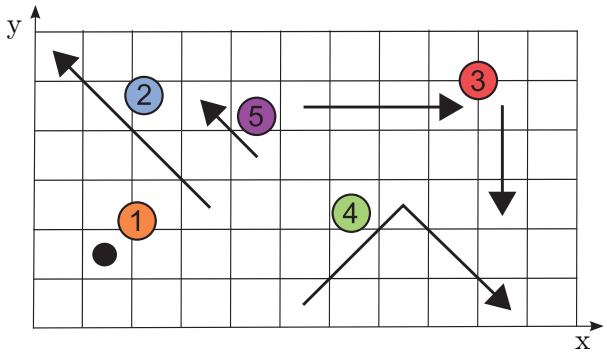

Figure 2: A receptor is moving in 2D space (velocity is constant along the depicted arrows). Different failure cases for ADRE can occur: Receptor r_1 (orange) has no or a too small velocity for resolution enhancement, r_2 (blue) is too fast, for r_3 (red) the resolution enhancement is only along the horizontal or vertical axis. The movement of r_4 (green) is optimal in direction but physically implausible. r_5 (magenta) shows a desirable movement for apparent resolution enhancement.

5 Extended model

The residual error e from minimizing Equation (5) is an objective measure of the quality of the ADRE algorithm under the assumption of no aliasing in the original image and perfectly reconstructed receptor movements, resp. optical flow.

In order to improve the result of the ADRE we use the formulation of Templin *et al.* [5] and propose to change the input $\mathbf{I_H}$ in a video by a transformation $\mathbf{T}$:

$$\mathbf{W}\left(\mathbf{I_L} \right) - \mathbf{T}(\mathbf{I_H}) = 0 \quad . \tag{6}$$

Note that in this case $\mathbf{W}$ and $\mathbf{T}$ are dependent variables because a change of $\mathbf{T}$ changes the optical flow and therefore $\mathbf{W}$. Restricting $\mathbf{T}$ to a discrete translation for each frame prevents resampling of the image which would otherwise annihilate most of the ADRE effect and render it useless. We will use the operator T^k to describe the absolute translation of the k^{th} frame I_H^k of the input video. For simplicity of explanation we will use T^k as both the translation function or the corresponding displacement vector as appropriate.

6 Saliency Model

The ADRE algorithms are based on the assumption that the receptors of the eye follow the optical flow in the video. The movement of the human eye however has only two degrees of freedom. If the optical flow is non-constant

across the image, e.g. different foreground and background movement, the integration result of the receptors is not in accordance to the optimization in Equation (5). It further implicates that in some cases no sufficient translation $\mathbf{T}$ can be found, as the required changes may cancel each other out. Assuming the eye movement is known, it is a valid simplification to optimize $\mathbf{T}$ only for receptors of the fovea due to the rapid falloff in acuity apart the foveal region [22]. We propose to use image saliency to model eye fixations. In our two-component saliency model, we compute a saliency map S^i from both objective, automatic saliency measures, and eye tracking data from a user study for each frame I_H^i of $\mathbf{I_H}$. We tested different saliency metrics stand-alone but the results turned out to be insufficient compared to results including the eye tracking data. If the quality of future saliency algorithms increases, a purely software-based solution for our approach is certainly possible. However, optimizing automatic saliency generation is not the scope of our work.

6.1 Subjective Saliency

We used an EyeLink 1000 eye tracker from SR Research for the eye tracking experiments. While subjects watched the videos their gaze paths (relative position within the video) were recorded at 240 Hz. 17 subjects with an average age of 25 and normal or corrected-to-normal vision participated in the user study for saliency generation. Since we were interested in the salient regions in the videos, we only needed to analyse the fixation points of the subjects, consisting of an x-y coordinate on the screen and a duration in milliseconds.

The output from our conducted eye tracking experiments is a single position vector for each frame of an animation per participant. Our assumption is that in the limit, i.e. with an infinite number of participants watching the animation for the first time, the normalized sum of the eye tracking data is the true saliency function $\mathbb{S}$. Hence, the data from our eye tracking experiments is a sparse sampling of $\mathbb{S}$ and estimating $\mathbb{S}$ becomes a reconstruction problem. Due to our restrictions on $\mathbf{T}$, a simple and fast reconstruction by downsampling the input images by three octaves, smoothing with a Gaussian of standard deviation $\sigma = 10$ and normalizing the resulting saliency maps S_E of each frame proved sufficient. In our experiments we found that even less participants, around ten, would have been sufficient because variance in gaze among viewers was low.

6.2 Objective Saliency Features

We apply the approach by Cerf *et al.* [18] that uses a combination of low-level features and high-level semantics as a second saliency metric. The low-level features are based on contrast of color, intensity and orientations

in the image [15]. For the high-level semantics we employed face detection [16] and person detection [17] as humans usually attract the most visual attention. Although there are offline face detectors reaching higher face detection rates [23], the detector of Viola and Jones proved to work fast and sufficiently robust for our test scenes. The result is saved in an intermediate saliency map S_O^i.

In combination we compute the final saliency map S^i for each frame as a weighted average of measured saliency S_E^i, predicted saliency S_O^i and a constant λ_s by the equation

$$S^i = \left((1-\lambda_s) + \lambda_s \cdot \left(\alpha S_O^i + (1-\alpha) S_E^i\right)\right). \tag{7}$$

The constant $\lambda_\mathbf{s} \in [0,1]$ steers the influence of non-salient regions. This is important in scenes where the foreground is moving fast and suffers from motion blur but the background contains fine details. In our experiments we used $\lambda_\mathbf{s} = 0.25$. α should be chosen depending on the reliability of the measured saliency S_E. In our experiments we used $\alpha = 0.25$.

The saliency map is thresholded before being used in our optimization described in the next section. This approach delivered sufficient maps in all our test cases. However, this approach can be costly and invasive. We alternatively allow to load salience maps which can be created semi-automatically, e.g. using Adobe After Effect's Rotobrush™. A coarse estimate of S is generally sufficient for our algorithm.

7 Trajectory Optimization

We postulate that a *sufficient* condition for $\mathbf{T}$ to serve as a good transformation for ADRE is given by four essential conditions:

1. Similarity to the optimal input for ADRE which is a one pixel flow along one of the diagonals per frame of the 120 Hz high-resolution video [3];
2. Smoothness of change;
3. Proximity to the original footage;
4. Visibility of image regions steered by saliency.

An explanatory example of the optimization is given in Fig. 3.

The last condition simplifies the optimization under the assumption that the flow inside the salient regions does not diverge. Let $\mathbf{u}^k$ be the optical flow from the original video $\mathbf{I_H}$ of frame k to $k+1$. Instead of evaluating and optimizing for every pixel of the video we compute the mean flow μ weighted by the saliency:

$$\mu^k = \sum_{i,j} S^k(i,j)\mathbf{u}^k(i,j), \tag{8}$$

(a) (b)

Figure 3: Explanatory example of our goal. (a) In the salient region (face of the person) the initial optical flow shows low diagonal movement. (b) After transformation of the frame using our approach, the optical flow shows an improved movement for ADRE in the same area.

The cumuluative sum of μ describes the trajectory of the salient region in $\mathbf{I_H}$.

Let $\mathbf{v}^k$ be the synthetically added translation we search for from frame k to frame $k+1$. The following energy terms incorporate the frame dependent weights $w^k_{\text{vel}} \in [0,1]$, $w^k_{\text{smooth}} \in [0,1]$ and $w^k_{\text{imp}} \in [0,1]$, which can be adjusted by our user interface described in Section 8. We formulate condition 1 as the difference of the newly synthesized flow from the optimal diagonal movement $\mathbf{v}_{\text{opt}}$:

$$E_{\text{vel}} = \sum_{k=1}^{n} w^k_{\text{vel}} \, \|\mu^k + \mathbf{v}^k - \mathbf{v}^k_{\text{opt}}\|_2^2 \tag{9}$$

Further, we enforce smoothness of the synthetic trajectory (condition 2) by minimizing the norm of the numerical derivative of the artificial flow:

$$E_{\text{smooth}} = \sum_{k=1}^{n-1} w^k_{\text{smooth}} \, \|\mathbf{v}^{k+1} - \mathbf{v}^k\|_2^2 \tag{10}$$

Finally, we prevent the translated video from drifting too far out of the original viewport by defining an additional energy term E_{imp}. For this, we first derive a distance map D^k from the saliency maps S^k. Each pixel in D^k saves its distance to the closest salient pixel. We then transform D^k into an what we call *importance map* V^k by mapping it to the range $[0,1]$ using $V^k(i,j) = 1 - D^k(i,j)/\max(D^k) + \epsilon$, with $\epsilon > 0$. A visualized example of the importance maps is given in Fig. 1. E_{imp} assures that the salient region S^k of each video frame stays within the original viewport V after applying the translation T^k. Additionally, the term penalizes large translations in

general.

$$E_{\text{imp}} = \sum_{k=1}^{n} w_{\text{imp}}^{k} \, (\sum_{(i,j)} V^k(i,j) - \sum_{(i,j)\in(VP\cap T^k(VP))} V^k(i,j))), \qquad (11)$$

E_{imp} computes the sum of all values in V in the transformed video outside the original viewport VP.

The final energy term is a weighted sum of its subterms:

$$E = \alpha \, E_{\text{vel}} + \beta \, E_{\text{smooth}} + \gamma \, E_{\text{imp}}. \qquad (12)$$

The weighting factors α, β and γ are chosen suitable for the coarse scale of the individual energy terms. For our test scenes $\alpha = 10^3$, $\beta = 10^{-3}$ and $\gamma = 10^4$ set the energy terms on equal ground.

We use the above formulation to perform an Expectation-Maximization-like optimization by iteratively refining an initial zero-vector $\mathbf{v} \in \mathbb{R}^{n\times 2}$. In each iteration, we alternate between finding an optimal movement $\mathbf{v}_{\text{opt}}$ for each frame in Equation (9), and updating $\mathbf{v}$ to minimize Equation (12) - note that at this stage we treat $\mathbf{v}$ as a vector field. Since $\mathbf{v}$ changes in this step, the classification of $\mathbf{v}_{\text{opt}}$ for each frame may also change. Therefore, the two steps are repeated until convergence which is guaranteed as each step minimizes Equation (12) further. The mean flow μ in Equation (9) assures that the natural movement in the video is taken into account during the optimization.

In more detail, in the Expectation-step, we minimize Equation (12) with respect to $\mathbf{v}_{\text{opt}}$ we choose the optimal flow

$$\mathbf{v}_{\text{opt}}^{k} \in \left\{(1,1)^\top, (1,-1)^\top, (-1,1)^\top, (-1,-1)^\top\right\}$$

at each frame k in order to minimize Equation (9). To prevent jittering of $\mathbf{v}_{\text{opt}}$ we keep $\mathbf{v}_{\text{opt}}$ constant for c frames. We therefore successively create bundles of c frames and compute the mean direction of the salient region, i.e. $\frac{1}{c}\sum_{i=1}^{c}(\mu^{k+i} + \mathbf{v}^{k+i})$, and choose the closest $\mathbf{v}_{\text{opt}}$ to this direction out of the four possibilities. c is a user defined variable. To further enforce smoothness of the overall trajectory of the salient region, we apply an additional diffusion step to the calculated $\mathbf{v}_{\text{opt}}$ by averaging each $\mathbf{v}_{\text{opt}}^{k}$ with its neighbors and repeating the process for m iterations. Per default we set $m = 5$.

In the Maximization-step, we employ a non-smooth numerical optimization method to minimize Equation (12) w.r.t. $\mathbf{v}$ [24]. Recall from Definition (11) that E_{imp} is not smooth with respect to $\mathbf{v}$. For fast convergence of the optimization algorithm, the exact gradient is necessary. The algorithm usually converges after two to four EM-steps. An example is given in Fig. 4.

Finally we obtain $\mathbf{v}^k + \mathbf{u}^k$ as the new optical flow.

7.1 Temporal upsampling

Due to the critical flicker frequency best temporal contrast is perceived for 120 Hz animations [3, 4]. Unfortunately, standard movies are captured at a much lower frame rate, usually 24 or 30 Hz. Simple duplication of the video frames is no recommendable solution as the flow between these will be zero in which case the ADRE algorithm cannot produce an enhanced output. Image interpolation algorithms would require a prohibitive resampling. To compensate for this, we first compute the optimized trajectory based on the original video, but the magnitude of the optimal flow $\mathbf{v}_{\text{opt}}$ and the initialization of $\mathbf{v}$ is multiplied by M, which is four or five for 30 Hz and 24 Hz respectively. Each image is then duplicated $M-1$ times, and we translate the m-th entity $I^{k,m}$, $m \in \{0, \dots M-1\}$, of image I^k according to:

$$T^{k,m} = T^k + \text{round}\left(\frac{m}{M}(\mathbf{v}^k + \mu^k)\right). \tag{13}$$

Not only does this enable ADRE as it smoothes the movement of the salient region for the duplicated frames, we also know the exact ground truth flow for each duplicated image from Equation (13).

Computing the correct optical flow for the original images can be delicate in very complex scenes. A wrong flow can result in severe artifacts when using ADRE. If we notice from the results that no sufficiently correct optical flow can be computed, we instead rely only on the displacement from Equation (13) as the input to the ADRE algorithm by Templin *et al.* [5].

8 User Interface Layout

We provide a user interface that lets the editor interactively modify the video trajectory by steering the optimization. This becomes necessary if the computed trajectory violates artistic intentions. A screenshot is shown in Fig. 5.

The components of the interface consist of a video preview window and an editing area. The video view shows the transformed video itself, the saliency map as an optional overlay (shown in yellow) and the action-safe frame (yellow) and title-safe frame (red), which both are standardized values in movie production. Using this view the editor can easily analyze the synthetic motion and its magnitude. The video is cut into single shots, which are optimized independently to avoid jittering artifacts at shot boundaries.

The editing window is subdivided into a navigation pane for playback and several panels below, which can be blended in or out as desired. Each one is either subject to one of the three energy terms in our energy minimization or one of the parameters influencing the optimization. At the top of each panel the user can specify keyframe values to set the relative influence of each parameter and energy weighting term throughout the video.

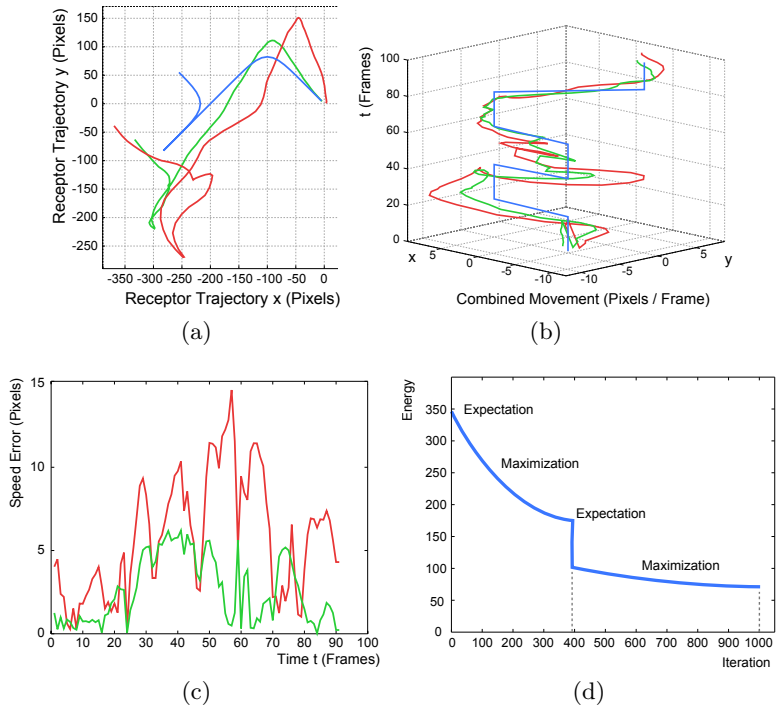

Figure 4: Analysis of our trajectory optimization for the scene EXPRESS: Visualization of the cumulated velocities (a) and velocities (b) as a function of space (x,y displacement) and time (frames). The optimized trajectory (green) still resembles the original trajectory (red) but is also closer to $\mathbf{v}_{\text{opt}}$ (blue) for optimized ADRE support. (c) The magnitude of the optimal velocity $\mathbf{v}_{\text{opt}}$ in the high-resolution 24 Hz video is $\sqrt{5^2+5^2}$ pixels. The red curve describes the deviation of the mean velocity μ from $\mathbf{v}_{\text{opt}}$ in the original video. The green curve describes the deviation after applying our optimization. Strong peaks give evidence of a change in the trajectory. (d) Plot of the energy level as a function of the number of maximization steps.

To the left of each of the error panels the weighting factor with regard to Equation (12) is set. All time-dependent parameters are linearly interpolated between each pair of keyframes. At the bottom the relative error of each energy function is plotted with a rainbow color scale. This gives the user a direct visual feedback on how his changes influence the quality of the later ADRE. Finally, the velocity control pane additionally contains three plots of the original velocity of the importance region (blue line), the optimized velocity (green line) and the theoretically optimal velocity (red dashed line). For example, in case a certain camera movement is essential in parts of the video the user simply increases the influence of the importance map term ('Visibility Control') for these frames and the optimizer adjusts the trajectory accordingly to follow the original motion path.

Figure 5: Screenshot of our interactive editor. Top right: A preview of the modified video. The video can be blended with the saliency maps (yellow). Bottom: Timeline of the video with a visualization of the different error terms. The user can interactively control the induced motion by setting keyframes to steer the influence of each term throughout the video. Error plots and trajectory for a new configuration are updated within seconds.

9 Experiments and Results

Our trajectory optimization implemented in MATLAB and C++ converges in one to five seconds for a 30 seconds, 24 Hz, 4K video, enabling interactive adjustment of the trajectory. However, considering the complete downsampling process, the computational complexity of our proposed technique is dominated by the ADRE algorithm by Templin *et al.* . Although [5] propose a fast GPU variant, we need to rely on a multi-threaded CPU version because of the sheer size of our video footage which did not fit into GPU memory. Therefore, computation takes around 30–45 seconds per subframe at 4K resolution on an Intel i7-960 with 3.2 GHz and 12 GB of RAM. For quality analysis of our work we created two video data sets where different goals have been pursued. The first data set (Maximum ADRE) containing 12 videos has been optimized for best ADRE effect and lower ranked similarity to the original video. For the second video data set (Minimum change) with eight video shots we optimized for a closer proximity of the synthetic trajectory to the original movement rather than optimal velocity and direction.

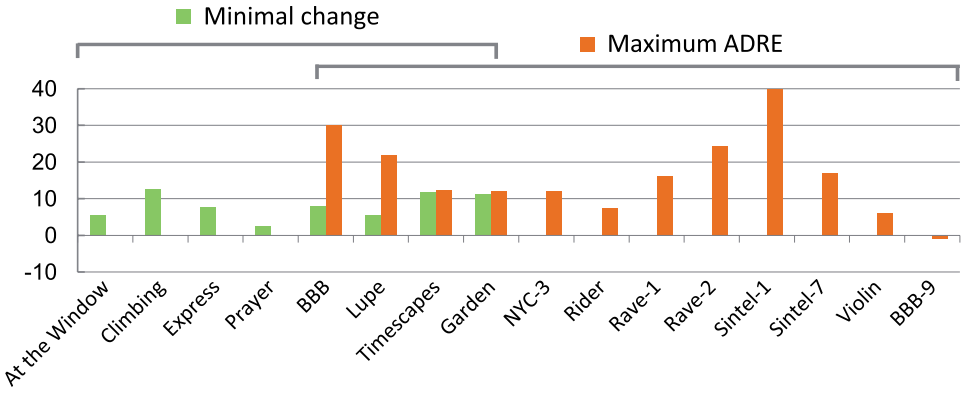

Figure 6: Improvement of the root mean squared error using our optimized variant (ours) as input to Templin's algorithm in comparison to using only the original video (default). Higher percentage values are better. One dataset (orange) is optimized for best ADRE effect whereby with the second (green) a trade-off between ADRE effect and low artificial movement is targeted.

9.1 Objective Enhancement – Statistics

The residuum of Equation 5 is an objective quality measure of the downsampling procedure assuming perfect SPEM. We compared the results using the original video and our optimized video as input to the algorithm of Templin *et al.* [5]. We achieve an improvement of the root mean square error (RMSE) which computes the difference of the high resolution frames and the perceived downsampled images of 17% on average for the Maximum ADRE data set and 8 % for the Minimum change data set, see Fig. 6. Videos of the second data set show less improvement for ADRE since the manipulated motion is kept close to the original one. Note that the overall RMSE can also increase in non-salient regions, e.g. in the Big Buck Bunny (BBB) scene, as we concentrated on locally optimizing for the salient regions. However, this happened only in a single scene and the increase in RMSE was below 1%.

9.2 Subjective Enhancement - User Study

To further validate the effectiveness of our approach we conducted a user study with 21 participants for both video data sets. All participants had normal or corrected-to-normal vision. They had no previous knowledge about the goal of the project nor the technique used. The subjects were seated in front of the monitor in a semidark room. They had been instructed orally regarding the procedure of the experiment. There was no time limit for solving the tasks.

Our aim was to show that our method outperforms previous downsam-

Figure 7: Example frames of our test scenes. Shown here are LUPE, TIMESCAPES, GARDEN and BIG BUCK BUNNY.

pling approaches for typical 24–30 Hz videos and that the editing tool enables to control the trade-off between noticeability of our applied transformation and effectiveness of our approach. We compared our method to Lanczos downsampling and the apparent display resolution enhancement of Templin *et al.* [5].

We used a 23.6 inch (diagonal) 120 Hz Acer GD245HQ display at a resolution of 1920×1080 pixels. The subjects viewed the monitor orthogonally at a distance of 10–30 inches. A 120 Hz frame refresh rate was considered, duplicating frames of the original video for Lanczos and Templin's algorithm. Hence, the optical flow was considered to be zero for all pixels between the duplicates. Note that using the original video for Templin's algorithm is not recommendable, as the integration time is too long with standard videos.

In our study we considered several test scenes containing varying types of motions (weak/strong), styles (real-world footage and animations), ambiguities in motion and saliency, and transparency (which is a yet unsolved challenge for optical flow in natural scenes). Example frames are shown in Fig. 7.

Motion perception In the first part of the study we analyzed the noticeability of our changes. We presented our modified 120 Hz videos to the first-time viewers without instructing them about our technique. The participants were then asked whether they noticed anything unnatural in the videos and if so what it was.

The analysis has shown that less than a third of the users noticed an obvious modification in the videos optimized for ADRE (Fig. 8, right). In these cases the original video material hardly contained any motion as a stylistic intent (SINTEL-7, VIOLIN, BBB-9). In all other cases our optimization was

able to produce a trajectory that was unsuspicious for first-time viewers. For the second data set manipulation was subtle so that almost no subject noticed a modification (Fig. 8, left).

Richness of Detail In the second part of the study we compared our results with Lanczos filtered videos (with lobe parameters 4,5 and 6) and Templin's ADRE [5] applied to the original 120 Hz video. The required optical flow for Templin was computed using the algorithm of Werlberger *et al.* [25] with an illumination corrected intensity term.

In pairwise comparisons the participants had to decide which video preserved more details, results are given in Fig. 9. The videos were shown in a loop. We randomly showed our approach on either the left or right side. When optimized for minimum change of the original camera path compared to Lanczos and Templin our approach is rated slightly better in terms of detail reconstruction for most scenes. However, significance is given only for three videos (Timescapes, Express, Prayer) for which the optimal movement can be achieved. For the videos of the second data set participants judged our technique significantly better. The results show that the degree of permitted manipulation strongly affects the perceivable improvement in apparent resolution enhancement. Compared to Templin we performed better in videos which originally contained little motion or appropriate motion magnitude but a non-diagonal direction of motion (Garden, Lupe, Prayer, BBB). We statistically validated this significance using a χ^2 test, successfully falsifying the null hypothesis. Thereby proving that there is a noticeable quality improvement with our approach in most cases. Our algorithm was even judged better in scenes where the RMSE globally increased (BBB-9).

Proximity to original movement The results show that the movements are noticable in case of adding motion to still shots when optimizing for maximum ADRE. However, the movements have not been rated disturbing. This shows that our tool enabled us to successfully avoid unnatural motions, especially in cases where the existing camera motion was of great aesthetic importance.

10 Discussion

Based on the findings from our user study on richness of detail we can infer that in general our method is able to achieve a statistically relevant improvement of the video quality over all others considered in this study. The smoothness factor also plays an important role in the perceived quality. Rough changes in the trajectory and especially possible jitter are strong perceptual artifacts.

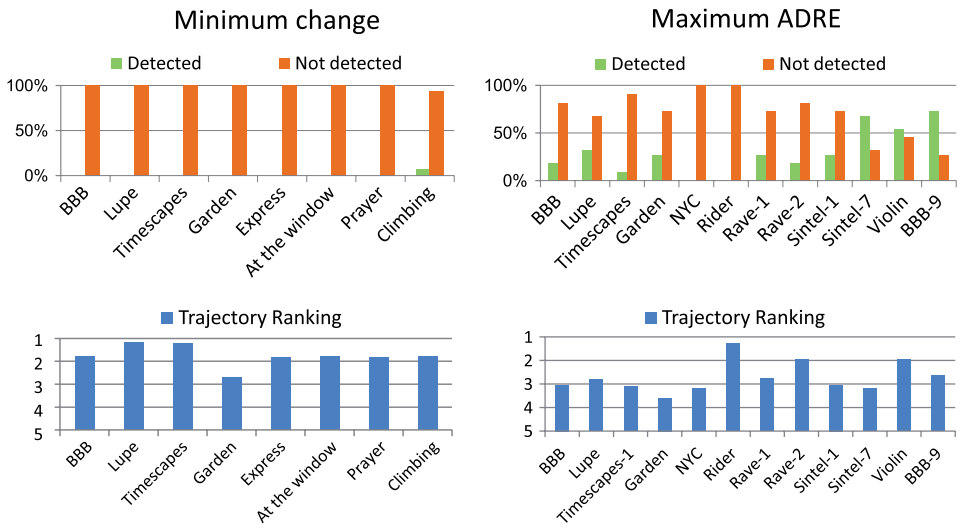

Figure 8: Trajectory conspicuity. (top) After viewing each video for the first time, participants stated if they had detected anything conspicuous regarding the video. (bottom) After viewing the video several times they were asked to rank the annoyance of the trajectory between 1 (pleasing), 3 (neutral) and 5 (disturbing).

Interestingly, most participants did not find out about our synthetically added trajectory if not told beforehand. The new trajectory resembles the movement of a steadicam or hand-held camera. Free camera motions filmed with Steadicams™ have become a popular visual style in many movies for which our approach can be directly applied and works best because the added motion is not noticed.

In our current implementation, the artificial trajectory applied to the input video results in undefined areas as we crop the video to the original viewport. In our test scenes with 4K resolution the amount of lost area always stayed below 4.7% of the full frame. Therefore, the visible area is always above the "action-safe" area of 95% and significantly above the "title-safe" area of 90%. For FullHD videos the value has been higher (up to 22.6%) to reach the optimal velocity for ADRE. However, the artist can directly control the amount of lost area for each part of the video by manipulating the importance map and adjusting the time-dependend weights of the importance term. Since undefined areas at the frame borders are seldomly in the focus of the viewer, they are unlikely to attract attention, especially when shown on large screens. Therefore, simple temporal inpainting techniques should be sufficient [26]. Furthermore, framing is common practice whenever video footage is shown on displays with different aspect ratios. We can optimize the cropping window as our algorithm automatically provides preservation of the salient region, smoothes the movement

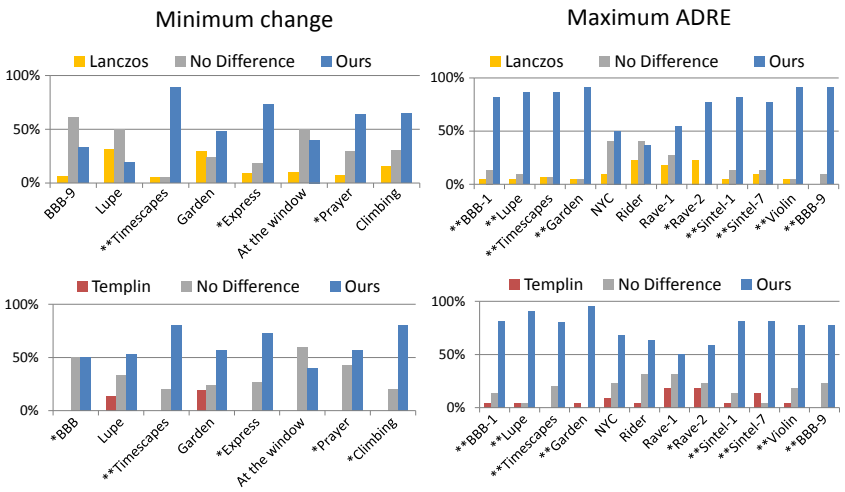

Figure 9: Our method compared to Lanczos and Templin when the participants watched the video in a loop. Significant differences according to the χ^2 test are marked with one or two asterisks, according to a significance level of 95% or 99%. We tested the Lanczos filter with lobe parameters 4, 5 and 6 but only listed the cases with the respective highest rating for Lanczos. Videos were either optimized for minimum change (left) or for highest ADRE (right).

of the window and pays attention to an optimal velocity for ADRE. The cropped area from framing can be used to fill in blank spaces that arise from our approach so that inpainting can be avoided. This problem, however, disappears if a slightly higher field-of-view is chosen during capturing of a scene or rendering of an animation.

The motion magnitude in our approach is dependent on the video content as well as the resolution of the original video. In most of our tests the high-resolution image was either of the size 4096×2304 pixels (4K) or 1920×1080 pixels (FullHD). An increased resolution naturally requires less noticeable synthetic changes in the video trajectory in most cases and hardly changes the intended overall composition of the video. As the market share of $2k$ projectors in cinemas is over 98%, our technique could be a valuable solution to increase the perceived resolution. The same holds for home cinemas.

One current limitation is the inability to faithfully enhance shots including large semi-transparent objects using default optical flow algorithms. The reason for this lies in the inability of the applied optical flow technique to distinguish between the different layers. It might be disputed that our underlying assumption of only one salient region in the image restricts us to only a subset of the possible shots in movies. However, one should keep in mind that fine details in movies, like pores or hair strands, are usually only visible in close-up views where our assumption is valid in most of the cases.

11 Conclusion

The gap in resolution between capturing and display devices requires downsampling of the original footage resulting in loss of fine details. Our approach provides an important step towards the preservation of these fine structures in standard videos. By moving the image along a smooth synthetic trajectory in combination with our temporal upsampling scheme a video can be optimized for apparent resolution enhancement. We evaluated benefits as well limitations of our approach in a perceptual study that shows apparent resolution enhancement is achieved even for complex scenes where previous approaches fail.

Still there is a variety of avenues for future research extending our proof of concept. The biggest challenge to bring the ADRE technique to market maturity is to create an appropriate compression scheme of the subsampled videos. A thirty seconds full HD video with 120 Hz has an uncompressed size of more than twenty gigabytes. Encoding the subframes directly is problematic as subframes generated by ADRE exhibit a lot of high-frequency details. Unfortunately, established video compression techniques rely on the assumption that large parts of the video can be predicted by displacing previous and future frames which does not hold in our case. Running Super-Resolution on previously downsampled videos as done by [19] would reduce the runtime for ADRE, but could also lead to a loss in detail quality and to additional artifacts as mentioned by the authors. Our approach enabling ADRE in arbitrary videos is fully compatible with this idea. A promising direction for saving bandwidth would be to compute the subframes in real-time. The overhead for our technique would be minimal, as only the 2D trajectory, two floating point values per frame, needs to be saved in addition to the video.

As stated in Section 10, ADRE for scenes containing semi-transparent objects is difficult as current optical flow algorithms in general assume opaque objects. To support such scenes a separation into different layers using matting algorithms and tracking of the separate layers is required.

Extending our approach to multiple saliency regions with conflicting flows would be possible by treating each region separately and deforming the rest of the image. If the warping is small enough it should not attract attention. Such a deformation was already successfully used for movie reshaping [27].

Although we concentrated on enhancing the salient regions in movies, it is possible to enhance manually specified parts of the video by adjusting the importance map. This could be interesting for sports events, e.g., to sharpen the advertisements in the background.

Acknowledgments

The authors would like to thank Ariana Prekazi for her great support in conducting the user study. We would also like to thank Krzysztof Templin, Stephen Higgins, Evin Grant, Keefa Chan and the Blender Foundation for offering usage of their work in our study.

References

[1] D. Purves, A. Shimpi, and B. R. Lotto, "An empirical explanation of the cornsweet effect," *J. Neuroscience*, vol. 19, no. 19, pp. 8542–8551, 1999.

[2] A. Yoshida, M. Ihrke, R. Mantiuk, and H.-P. Seidel, "Brightness of the glare illusion," in *Proc. of the Symposium on Applied Perception in Graphics and Visualization*, pp. 83–90, ACM, 2008.

[3] P. Didyk, E. Eisemann, T. Ritschel, K. Myszkowski, and H. Seidel, "Apparent display resolution enhancement for moving images," *ACM Transactions on Graphics*, vol. 29, no. 4, pp. 113:1–113:8, 2010.

[4] M. Kalloniatis and C. Luu, "Temporal resolution," 2007. http://webvision.med.utah.edu/temporal.html, visited Sep. 2012.

[5] K. Templin, P. Didyk, T. Ritschel, E. Eisemann, K. Myszkowski, and H.-P. Seidel, "Apparent resolution enhancement for animations," in *Proc. of Spring Conference on Computer Graphics*, pp. 85–92, 2011.

[6] F. Berthouzoz and R. Fattal, "Apparent resolution enhancement for motion videos," in *Proceedings of the ACM Symposium on Applied Perception*, pp. 91–98, ACM, 2012.

[7] D. P. Mitchell and A. N. Netravali, "Reconstruction Filters in Computer-Graphics," in *Conference on Computer graphics and interactive techniques*, ACM Siggraph, pp. 221–228, 1988.

[8] F. Berthouzoz and R. Fattal, "Resolution enhancement by vibrating displays," *ACM Transactions on Graphics*, vol. 31, no. 2, pp. 15:1–15:14, 2012.

[9] N. Damera-Venkata and N. Chang, "Display supersampling," *ACM Transactions on Graphics*, vol. 28, no. 1, pp. 9:1–9:19, 2009.

[10] W. Allen and R. Ulichney, "Wobulation: Doubling the addressed resolution of projection displays," in *Proc. of the International Symposium Digest of Technical Papers*, vol. 47, pp. 1514–1517, Society for Information Display, 2005.

[11] Z. Hara and N. Shiramatsu, "Improvement in the picture quality of moving pictures for matrix displays," *Journal of the Society for Information Display*, vol. 8, no. 2, pp. 129–137, 2000.

[12] J. Platt, "Optimal filtering for patterned displays," *Signal Processing Letters, IEEE*, vol. 7, no. 7, pp. 179–181, 2000.

[13] D. S. Messing and L. J. Kerofsky, "Using optimal rendering to visually mask defective subpixels," in *Human Vision and Electronic Imaging XI*, SPIE Proceedings Series, pp. 236–247, 2006.

[14] S. Basu and P. Baudisch, "System and process for increasing the apparent resolution of a display," 2009. US Patent 7,548,662.

[15] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 11, pp. 1254–1259, 1998.

[16] P. Viola and M. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[17] P. Felzenszwalb, D. Mcallester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[18] M. Cerf, J. Harel, A. Huth, W. Einhäuser, and C. Koch, "Decoding what people see from where they look: Predicting visual stimuli from scanpaths," *Attention in Cognitive Systems*, pp. 15–26, 2009.

[19] G. Boccignone, A. Marcelli, and G. Somma, "Analysis of dynamic scenes based on visual attention," *Proceedings of AIIA*, 2002.

[20] H. van Hateren, "A cellular and molecular model of response kinetics and adaptation in primate cones and horizontal cells.," *Journal of vision*, vol. 5, no. 4, pp. 331–347, 2005.

[21] J. Lairda, M. Rosen, J. Pelz, E. Montag, and S. Daly, "Spatio-Velocity CSF as a function of retinal velocity using unstabilized stimuli," in *Human Vision and Electronic Imaging XI*, vol. 6057 of *SPIE Proceedings Series*, pp. 32–43, 2006.

[22] C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson, "Human photoreceptor topography," *Journal of Comparative Neurology*, vol. 292, no. 4, pp. 497–523, 1990.

[23] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2879–2886, IEEE, 2012.

[24] M. Overton, "HANSO: Hybrid Algorithm for Non-Smooth Optimization 2.0," 2010. http://www.cs.nyu.edu/overton/software/hanso/.

[25] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof, "Anisotropic Huber-L^1 optical flow," in *Proc. of the British Machine Vision Conference*, pp. 1–11, 2009.

[26] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '00, pp. 417–424, 2000.

[27] A. Jain, T. Thormählen, H.-P. Seidel, and C. Theobalt, "MovieReshape: tracking and reshaping of humans in videos," *ACM Trans. Graph.*, vol. 29, pp. 148:1–148:10, 2010.

[28] O. Artamonov, "X-bit's guide: Contemporary lcd monitor parameters and characteristics," 2004. `http://www.xbitlabs.com/articles/monitors/display/lcd-guide_11.html`, visited December 2011.

[29] M. Boehme, M. Dorr, C. Krause, T. Martinetz, and E. Barth, "Eye movement predictions on natural videos," *Neurocomputing*, vol. 69, no. 16-18, pp. 1996–2004, 2006.

[30] M. Cerf, J. Harel, W. Einhäuser, and C. Koch, "Predicting human gaze using low-level saliency combined with face detection," *Advances in neural information processing systems*, vol. 20, 2008.

[31] M. Dorr, T. Martinetz, K. Gegenfurtner, and E. Barth, "Variability of eye movements when viewing dynamic natural scenes," *Journal of Vision*, vol. 10, no. 10, 2010.

[32] M. Guttmann, L. Wolf, and D. Cohen-Or, "Content aware video manipulation," *Computer Vision and Image Understanding*, 2011.

[33] A. Lewis and M. Overton, "Nonsmooth optimization via bfgs," *SIAM Journal of Optimization*, pp. 1–35, 2009.

[34] Z. Hara and N. Shiramatsu, "Improvement in the picture quality of moving pictures for matrix displays," vol. 8, no. 2, pp. 129–137, 2000.

[35] Z. Hara, N. Terazaki, N. Shiramatsu, and S. Iwata, "Picture quality of different pixel arrangements for large-sized matrix displays," *Electronics and Communications in Japan (Part II: Electronics)*, vol. 77, no. 7, pp. 105–120, 1994.

[36] Y. Hu, D. Rajan, and L. Chia, "Adaptive local context suppression of multiple cues for salient visual attention detection," in *Multimedia and*

Expo, 2005. ICME 2005. IEEE International Conference on, pp. 4–pp, IEEE, 2005.

[37] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 2106–2113, IEEE, 2009.

[38] C. Kanan and G. Cottrell, "Robust classification of objects, faces, and flowers using natural image statistics," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 2472–2479, IEEE, 2010.

[39] J. Kucerova, "Saliency map augmentation with facial detection,"

[40] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or, "Optimizing photo composition," in *Computer Graphics Forum*, vol. 29, pp. 469–478, Wiley Online Library, 2010.

[41] T. Liu, N. Zheng, W. Ding, and Z. Yuan, "Video attention: Learning to detect a salient object sequence," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4, IEEE, 2008.

[42] J. Napoli, S. Dey, S. Stutsman, O. Cossairt, T. Purtell II, S. Hill, and G. Favalora, "Imaging artifact precompensation for spatially multiplexed 3-d displays," *Proc. of SPIE*, 2008.

[43] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 3^{rd} ed., 2007.

[44] J. Van Ouwerkerk, "Image super-resolution survey," *Image and Vision Computing*, vol. 24, no. 10, pp. 1039–1052, 2006.

[45] M. Werlberger, T. Pock, and H. Bischof, "Motion estimation with non-local total variation regularization," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2010.

[46] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–6, IEEE, 2007.

[47] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th annual ACM international conference on Multimedia*, pp. 815–824, ACM, 2006.

Michael Stengel received his Diploma degree in Computational Visualistics from the University of Magdeburg, Germany, in 2011. From 2010 to 2011 he worked at the Virtual Reality Lab at Volkswagen AG. He is currently pursuing his PhD in Computer Graphics at Technische Universität (TU) Braunschweig, Germany. His research interests include visual perception, human-computer-interaction and visualization.

Martin Eisemann received a Diploma degree in Computational Visualistics from the University of Koblenz-Landau, Germany, in 2006 and his PhD degree in Computer Graphics from the TU Braunschweig, Germany, and received the best student paper award at the annual conference of the European Association for Computer Graphics (Eurographics) in 2008. Since 2011 he is Akademischer Rat (Postdoctoral Researcher) at the Computer Graphics Lab at the TU Braunschweig. His main research interests include image- and video-based rendering and editing, visual analytics, and realistic and interactive rendering.

Stephan Wenger received his Diploma degree in Computer Science in 2009 and his Diploma in Physics in 2010 from Technische Universität (TU) Braunschweig, Germany. He is currently pursuing his PhD in Computer Science at the Institute for Computer Graphics at TU Braunschweig. His research interests include 3D reconstruction and visualization, astronomical phenomena, as well as audio and music processing.

Benjamin Hell received his Diploma degree in Industrial Mathematics in 2011 from University of Bayreuth, Germany. He is currently pursuing his PhD in Applied Mathematics at the Institute for Computer Graphics at TU Braunschweig. His research interests include optimization and geometry.

Marcus Magnor is full professor and head of the Computer Graphics Lab at Braunschweig University of Technology. He holds a Diploma degree in Physics (1997) and a PhD in Electrical Engineering (2000). After his post-graduate time as Research Associate in the Graphics Lab at Stanford University, he established his own research group at the Max-Planck-Institut Informatik in Saarbrücken. He completed his habilitation and received the venia legendi in Computer Science from Saarland University in 2005. In 2009, he spent one semester as Fulbright scholar and Visiting Associate Professor at the University of New Mexico. His research interests meander along the visual information processing pipeline, from image formation, acquisition, and analysis to image synthesis, display, perception, and cognition. Ongoing research topics include image-based measuring and modeling, photorealistic and real-time rendering, and perception in graphics.